Understanding Respondents' Attitudes Toward Web Paradata Use

Kunz, Tanja; Gummer, Tobias

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Kunz, T., & Gummer, T. (2020). Understanding Respondents' Attitudes Toward Web Paradata Use. *Social Science Computer Review*, *38*(6), 739-753. https://doi.org/10.1177/0894439319826904

Article

Social Science Computer Review
2020, Vol. 38(6) 739-753

DOI: 10.1177/0894439319826904
journals.sagepub.com/home/ssc
SAGE

Understanding Respondents' Attitudes Toward Web Paradata Use

Tanja Kunz[1] and Tobias Gummer[1]

Abstract

The collection and use of paradata is gaining in importance, especially in web surveys. From a research ethics' perspective, respondents should be asked for their consent to the collection and use of web paradata. In this context, a positive attitude toward paradata use has been deemed to be a prerequisite for respondents' willingness to share their paradata. The present study aimed to identify factors affecting respondents' attitudes toward paradata use. Our findings revealed that adequately informing survey respondents about what paradata are and why they are used was an important determinant of their attitudes toward paradata use. Moreover, we found that respondents with a positive attitude toward the survey were more likely to have a favorable opinion of paradata use. Our findings suggest that a thorough understanding of the factors that contribute to a positive attitude toward paradata use provides the basis for improved paradata consent procedures, which in turn will increase rates of consent to paradata use and help attenuate the risk of consent bias in web surveys.

Keywords

paradata, respondents' attitudes, informed consent, consent bias, web survey

The collection and use of paradata in web surveys has increased continuously in recent years—with the primary focus being on paradata that can be used to address measurement issues. The term *paradata* refers to information about the data collection process that is gathered as a by-product of computer-assisted data collection (Couper, 1998; Kreuter, 2013). Two main classes of paradata can be distinguished in web surveys: device-type paradata and questionnaire navigation paradata (Callegaro, 2013). Device-type paradata (e.g., information about the browser used, the operating system, and the device type) are primarily used to customize questionnaires to specific browsers and devices. Questionnaire navigation paradata (e.g., time stamps, mouse clicks, and loss of focus) can be used to reconstruct and describe the questionnaire completion process and allow for conclusions to be drawn about the survey-taking experience. Both device-type and questionnaire navigation paradata can be

[1] GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany

Corresponding Author:
Tanja Kunz, GESIS—Leibniz Institute for the Social Sciences, P. O. Box 12 21 55, Mannheim 68072, Germany.
Email: tanja.kunz@gesis.org

used to evaluate and improve survey instruments, to better understand how respondents answer questions, and, ultimately, to improve the quality of measurement in web surveys (Callegaro, 2013; McClain et al., 2019).

Given the increasing use of web paradata, researchers must decide whether and how respondents should be informed about the capture and use of the paradata that are collected while they are filling out the survey (Couper & Singer, 2013). Guidelines addressing legal, ethical, and practical considerations that survey researchers should take into account when conducting public opinion, market, or social research have been published by professional associations in these fields (e.g., ESOMAR, CASRO, and AAPOR). However, these guidelines remain vague with respect to standards for seeking respondents' consent to the capture and use of paradata. Thus, existing recommendations on informed consent are often open to interpretation. For instance, the American Sociological Association (2018) Code of Ethics sets forth that consent from research participants should be obtained "(1) when data are collected from research participants through any form of communication, interaction, or intervention; or (2) when behavior of research participants occurs in a private context where an individual can reasonably expect that no observation or reporting is taking place" (pp. 12, 13). In Germany, the joint guideline for online surveys published by the professional associations ADM, ASI, BVM, and DGOF (2007) recommends that, in cases where additional data are to be imperceptibly collected and stored during a survey, respondents should be requested to give their prior express consent thereto. Whether this recommendation relates also to paradata, and, if so, whether all types of paradata are included, is open to interpretation. Existing guidelines are clearer when it comes to personal data. However, a clear distinction cannot always be made between paradata and personal data (also called personally identifiable information) collected during web surveys. "Personal data" refer to information that can be used to identify a person directly or indirectly "by reference to an identification number or the person's physical, physiological, mental, economic, cultural or social characteristics" (ESOMAR/GRBN, 2015, p. 8). For instance, geolocation data (latitude and longitude) captured in web surveys serve to identify the geographic location of the computing device (desktop computer, tablet, smartphone, etc.) used to complete the survey. These data can therefore be considered to be paradata, and respondents must be informed about and asked for their permission for their collection and use (ESOMAR/GRBN, 2015). Although in many instances still a matter of interpretation, the ethical issues involved in collecting paradata can be expected to continue gaining importance. This is because, on the one hand, the relevant data protection regulations are subject to an ongoing process of change, and, on the other hand, the possibilities for collecting even more detailed paradata continue to grow. This is especially evident in web surveys, due to the availability, accessibility, and continuous development of paradata scripts that allow for the collection of detailed client-side paradata (Heerwegh, 2003; Kaczmirek & Neubarth, 2007; Schlosser & Höhne, 2018).

Once the decision has been made to inform respondents about the capture and use of paradata in a web survey, the question arises as to how best to inform them. To date, little research has been conducted on this issue. Survey researchers can implicitly seek consent by merely informing respondents about their intention to capture and use paradata, or they can explicitly seek consent by directly asking respondents for their permission to do so. In either case, this is likely to have negative effects on respondent behavior. In the worst case, individuals will refuse to take part in the web survey solely because the capture and use of paradata have been mentioned. As a consequence, no survey data and paradata would be available for these nonrespondents. If individuals agree to participate in the survey but refuse to consent to the collection and use of paradata, the linking of survey data and paradata will not be possible in their case, and the sample size for analyses that draw on both types of data will be reduced. Moreover, consent bias is likely to be introduced when the group of respondents who consented to the collection and use of paradata differs systematically from the group who declined to consent (Couper & Singer, 2013; Revilla, Couper, & Ochoa, 2019).

To gain a better understanding of how best to obtain respondents' consent for paradata collection and use, and how to avoid biases associated with this process, we need to better understand the factors influencing respondents' consent behavior. According to the reasoned action framework proposed by Fishbein and Ajzen (2010), people's attitudes toward a behavior are a crucial determinant of their behavioral intentions and behaviors. In other words, the more favorable people's attitudes toward a behavior are, the stronger their intention to perform it will be. People's attitudes toward personally performing a behavior are, in turn, determined by beliefs about its positive and negative consequences. Among "the multitude of variables [that] could potentially influence the beliefs people hold" (e.g., gender, age, education, and past experiences), Fishbein and Ajzen (2010, p. 24) included exposure to information, which is considered to be an influencing factor when balancing the positive and negative consequences of a behavior. If people expect that the performance of a behavior will result in more positive than negative outcomes, their attitudes toward that behavior will, in general, be favorable (Fishbein & Ajzen, 2010, p. 20). Consequently, attitudes toward a behavior can be considered to be the result of cost–benefit balancing, with a favorable attitude making performance of that behavior more likely.

Previous research has provided surprisingly little insight into respondents' attitudes toward paradata use and, thus, into the factors behind their willingness to share web paradata. Recent studies have focused primarily on respondents' actual or stated willingness to consent to paradata use, without adequately understanding respondents' motives for acceptance or refusal. From studies on respondents' consent to share different types of passive behavior tracking data (e.g., biomeasures, URLs of visited websites, GPS coordinates), we know that the willingness to share these data varies across tasks (i.e., the type of data) and across respondents. For instance, only 20.8% of the members of an opt-in panel in Spain stated that they would be willing to share GPS information from their smartphones. This was thus among the least accepted of a total of 20 different mobile data collection tasks (Revilla et al., 2019). Wenz, Jäckle, and Couper (2017) asked members of a probability-based online panel in the United Kingdom about their willingness to complete eight different mobile data collection tasks and found that only 39% of the smartphone users would have been "very" or "somewhat" willing to share the GPS position of their smartphone. Again, this was among the least accepted tasks in the survey. Overall, lower levels of willingness to perform different mobile data collection tasks were found for those tasks that posed a potential threat to the respondents' privacy. Moreover, respondents stated that they would be more willing to use their tablets than their smartphones to perform a task that posed a potential threat to privacy (i.e., downloading and using a tracking app that collected anonymous data about how they used the device; Wenz, Jäckle, & Couper, 2017). Toepoel and Lugtig (2014) asked members of a probability-based online panel in the Netherlands for their permission to record GPS coordinates. They found that about a quarter of respondents gave permission, with consent rates among smartphone users (26%) being similar to those among desktop users (24%). Especially due to the availability of mobile apps, smartphones have become an essential part of everyday life and are used for personal communication, shopping, navigation, entertainment, and so forth. Although personal data such as web use and browsing history, app usage statistics, and voice or audio recordings can be collected only via mobile apps, and although the collection and processing of these data is subject to the individual's explicit prior consent (ESOMAR/GRBN, 2017), it is conceivable that respondents consider consent to the use of paradata collected via their smartphones to be a greater privacy risk. This may be due to the fact that, when making their consent decision, respondents have the various types of personal data in mind that can potentially be collected on their smartphones.

Revilla, Couper, and Ochoa (2019) found that respondents with a positive attitude toward surveys (i.e., those who answered the income question, who reported that they liked answering the survey, and who had responded to a greater number of prior panel surveys) were more willing to share different types of passive behavior tracking data. Surprisingly, Wenz et al. (2017) found that

respondents who had been part of the panel for a longer time were less cooperative with additional passive data collection requests than those who had joined the panel more recently. The intensity of smartphone use (i.e., the number of different activities respondents performed on their smartphone; Wenz et al., 2017), the frequency of related smartphone use (e.g., the frequency of using GPS information for their own purposes), and previous participation in similar tasks (e.g., the frequency of sharing GPS information) also increased respondents' willingness to perform additional passive data collection tasks (Struminskaya et al., 2018). Respondents who had greater trust in the anonymity of these types of data were also more willing to share them, whereas particularly those respondents who raised concerns about privacy or confidentiality were less likely to consent (Couper & Singer, 2013; Revilla et al., 2019; Wenz et al., 2017). When respondents were asked for the reasons for their acceptance or refusal to share these data, privacy, security, and trust were among the most frequently cited reasons (Couper & Singer, 2013; Keusch, Antoun, Couper, Kreuter, & Struminskaya, 2017; Revilla et al., 2019), as was insufficient information about the passive data collection task (Keusch et al., 2017). In contrast, those respondents who saw benefits for themselves or society were found to be more willing to share this type of data (Revilla et al., 2019). In a wording experiment, Sakshaug and Kreuter (2014) found that respondents were more likely to consent to the linking of survey and administrative data when the linkage request contained a respondent-benefit wording (e.g., "to keep the interview as brief as possible") than when it was more neutrally worded. Thornby, Calderwood, Kotecha, Beninger, and Gaia (2018) provided respondents with eight different wordings decribing the benefits of administrative data linkage and asked them for the reasons why they liked or disliked the respective benefit wordings. Evidence from the qualitative interviews showed that wordings were preferred that stressed the usefulness of administrative data linkage and described the benefits for the individual participants in a simple and clear way. When asked what they did not like, respondents reported difficulties with wordings that addressed benefits that were not relevant to individual participants. In line with this finding, Couper and Singer (2013) found that mentioning benefits for the survey researchers had no beneficial effect on respondents' likelihood to consent to web paradata use. Thus, mentioning benefits for persons other than the respondents themselves does not necessarily outweigh the perceived costs of sharing web paradata. Overall, previous research has concentrated mainly on the actual or stated willingness to consent to administrative data linkage and on different passive data collection tasks on mobile devices without sufficiently investigating the factors that influence the consent decision.

The present study was aimed at addressing the research gap concerning respondents' attitudes toward web paradata use and the factors influencing these attitudes. We drew on the reasoned action framework proposed by Fishbein and Ajzen (2010) to better understand how exposure to information about paradata affects respondents' attitudes toward paradata use. Specifically, we wanted to answer two research questions. First, how do the provision of clarifying information and the mentioning of benefits affect respondents' attitudes toward web paradata use? And second, what factors contribute to a positive attitude toward web paradata use? To answer these research questions, we conducted a survey experiment in which we provided additional clarifying information about paradata and mentioned benefits of paradata use either for the researchers or the respondents. In the next section, we outline our research hypotheses. This is followed by a presentation of our data and methods. After discussing our results, we close with concluding remarks.

Hypotheses

In line with the reasoned action approach (Fishbein & Ajzen, 2010), we expect cost–benefit balancing to affect respondents' attitudes toward paradata use and, ultimately, to determine their willingness to consent to it. Moreover, it is reasonable to assume that beliefs about the positive and negative consequences of paradata sharing may vary depending on the paradata type in question

(e.g., geolocation data, keystrokes). As paradata types are expected to differ in their perceived level of sensitivity or intrusiveness, the costs that respondents associate with sharing different types of paradata may also vary, and this in turn would result in differences in the respondents' attitudes toward paradata use. These deliberations gave rise to the following hypotheses:

Hypothesis 1: Respondents who receive a short definition of the paradata to be collected will be better able to form an opinion on the use of these paradata than respondents who do not receive any additional information.

Hypothesis 2: Respondents who receive a short definition of the paradata to be collected will have a more positive attitude toward the use of these paradata.

Hypothesis 3: The positive effect on respondents' attitudes toward paradata use will be even more pronounced when specific benefits of the use of these paradata are made explicit to the respondents.

Hypothesis 4: Respondents' attitudes toward paradata use will be more positive when the benefits mentioned are relevant to the respondents rather than the researchers.

To better understand what determinants contribute to respondents' positive attitudes toward web paradata use, we examined various respondent-related characteristics. In addition to basic socio-demographic variables (i.e., gender, age, and education), we looked at different aspects of respondents' assessments of the survey (i.e., as interesting, varied, important for science, long, difficult, or too personal). With reference to the reasoned action framework, we made the following prediction:

Hypothesis 5: The more favorable the respondents' overall assessments of the survey are, the more likely they will be to be positively disposed toward paradata use.

And finally, the type of device used to take the survey was examined in order to test the following hypothesis:

Hypothesis 6: Smartphone respondents will be more negatively disposed toward paradata use than respondents who complete the survey on a desktop computer or a tablet.

Data and Method

Sample

Our experiment was embedded in a web survey conducted in November 2017 among members of a large German opt-in online panel provided by respondi AG. The target population comprised individuals aged 18–69 years living in households with Internet access in Germany. The sample was based on quotas for gender, age, and education. From the 3,030 panelists invited to participate in the survey, 498 were screened out and 2,247 completed the survey, resulting in a completion rate of 89% (Callegaro & DiSogra, 2008). The breakoff rate was 7%, which can be considered rather low, as it was less than half the average breakoff rate in web surveys of similar length, samples, and topic in Germany (16% based on 18 comparable web surveys; Gummer, Quoß, & Roßmann, 2019). Of those who completed the survey, 51% were female, the average age was 45 years, and 34% had a high level of education (university entrance qualification). The topic of the survey was "Politics and Work." The questionnaire included questions on voting, political knowledge, and policy preferences, which were adapted from the German Longitudinal Election Study; it also measured attitudes toward work, achievement motivation, and work–life balance. The questionnaire concluded with a question module on the respondent's assessment of the survey. On average, the questionnaire took

Table 1. Information on Paradata Provided to the Control Group and the Three Experimental Groups.

(a) No Definition/ No Benefit	(b) Definition/ No Benefit	(c) Definition/ Benefit for Researchers	(d) Definition/ Benefit for Respondents
(1) Type of device used	Information about whether a PC, tablet, or smartphone was used to answer the questions	Help to better understand the response behavior	Help to optimally display the questionnaire on your device
(2) Time stamps for every mouse click	Information about how long it takes to answer each question	Help to better assess the quality of survey data	Help to make it easier for you to complete the questionnaire
(3) Geolocation information	Information about the environment in which the survey has been completed	Help to better understand the survey situation	Help to choose survey topics that are interesting for you

32.9 min to complete ($Mdn = 29.3$). A responsive questionnaire design was used; 21% of the respondents answered the survey on a smartphone.

Experimental Design

The experimental question asked all respondents for their attitudes toward paradata use in web surveys. The question stem was worded as follows: "In addition to survey data, other so-called technical data can be collected during a survey. How positively or negatively disposed are you toward the use of the following technical data, exclusively for scientific purposes?" The experimental question was answered on 4-point scale with the labels *very positively*, *rather positively*, *rather negatively*, and *very negatively*. An explicit "don't know" option was provided. All answers were voluntary, and respondents were allowed to skip items without being prompted.

The experimental question comprised 3 items on different types of paradata that are frequently used for methodological purposes, namely (1) device type, (2) mouse click time stamps, and (3) geolocation information. All 3 items were presented together on one webpage. Device-type paradata provide information about the kind of device respondents use to complete a questionnaire. A lot of related research deals with tailoring question layouts for smartphones and examining how smartphone respondents differ from desktop respondents (for a comprehensive overview, see Couper, Antoun, & Mavletova, 2017). Client-side time stamps for each mouse click provide information on the time that elapses between each mouse click and can be used to identify question-level response times that are unusually short or long. This information can be used to identify difficult and burdensome questions (Kunz, 2015; Yan & Tourangeau, 2008) and to detect and prevent satisficing behaviors such as speeding, straightlining, and clicking through (Conrad, Tourangeau, Couper, & Zhang, 2017; Revilla & Ochoa, 2015; Stieger & Reips, 2010). Geolocation information (i.e., GPS coordinates) is used to determine where respondents are when they complete the survey. This is considered especially useful when conducting mobile surveys or collecting time diary and travel diary data (Mavoa, Oliver, Witten, & Badland, 2011; Toepoel & Lugtig, 2014).

Table 1 shows which versions of the experimental question were presented to the respondents in the different conditions. The question stem and the response scale were identical in all conditions, whereas the wording of the items differed across conditions. In a between-subjects design, respondents were randomly assigned to the control group or to one of the three experimental groups. The conditions differed in terms of whether a definition of the paradata in question was provided and whether a benefit of the use of these paradata was explicitly mentioned. In the control group (a), the paradata type was stated (e.g., "Type of device used"), but no definition was provided and no benefit of the use of these paradata was mentioned. In the first experimental group (b), a short definition of

Table 2. Attitudes Toward Paradata Use in the Control Group and the Three Experimental Groups (in Percent).

		(a) No Definition/ No Benefit	(b) Definition/ No Benefit	(c) Definition/ Researcher Benefit	(d) Definition/ Respondent Benefit	χ^2 Test	Cramer's *V*
(1) Device type	Positive	79.9[d]	84.1	79.3	86.0[a]	13.91(6)*	.056
	Negative	8.3	6.9	7.5	5.0		
	No opinion	11.8	9.0	13.3	9.0		
(2) Time stamps	Positive	65.2[b]	74.7[a]	69.9	71.5	17.70(6)**	.063
	Negative	16.8	14.2	16.1	17.0		
	No opinion	18.0[b]	11.1[a]	14.0	11.5		
(3) Geolocation	Positive	52.0[b]	61.6[a]	59.3	58.1	16.09(6)*	.061
	Negative	34.0	29.0	27.5	28.0		
	No opinion	14.0	9.4	13.3	14.0		
	n	577	521	535	565		

Note. $N = 2{,}198$; n = number of cases in each group. χ^2 test of independence with *df* in parentheses and pair-wise comparisons (Bonferroni correction), with superscripts indicating a significant difference ($p < .05$ or less) between any two of the four groups, that is, compared to no definition/no benefit (a), definition/no benefit (b), definition/researcher benefit (c), or definition/respondent benefit (d).
*$p < .05$. **$p < .01$. ***$p < .001$.

the paradata in question was provided, which briefly described the kind of information collected (e.g., "Type of device used: Information about whether a PC, tablet, or smartphone was used to answer the questions"), but no benefit was mentioned. In the second and third experimental groups (c) and (d), a short definition of the paradata in question was provided and a benefit was mentioned that was of relevance either to the survey researcher (experimental group [c]) or to the respondent (experimental group [d]). For example, in experimental group (c), the information provided was "Type of device used: Information about whether a PC, tablet, or smartphone was used to answer the questions helps to better understand the response behavior." (Figure A1 in the Appendix shows screenshots of each version.) Benefits for the researchers primarily related to the quality of survey response, whereas benefits for the respondents emphasized the ease and enjoyment of response to the present survey and—because respondents were active members of an online panel—presumably to future surveys in that panel as well.

The experimental question was embedded in a series of questions on the respondent's assessment of the survey, which were asked in the last quarter of the questionnaire. The survey assessment module began with two questions about different aspects of survey assessment and respondent effort, followed by the experimental question on the respondent's attitude toward paradata use in web surveys. After the experimental question, we asked respondents an open-ended question about potential reasons for not consenting to web paradata use. All respondents received this open-ended question, irrespective of whether they had indicated a positive or negative attitude toward paradata use.

Results

Table 2 presents the results of χ^2 tests of independence that were conducted to answer the question of how the provision of a short definition and the mention of a benefit affected respondents' attitudes toward paradata use in web surveys. Responses to the experimental question of how positively or negatively the respondent was disposed toward the use of different kinds of paradata were classified as positive ("very positively," "rather positively"), negative ("rather negatively," "very negatively"), or no opinion ("don't know"). Overall, our findings revealed a rather positive attitude toward web

paradata use for scientific purposes, as indicated by high proportions of positive attitudes toward the use of information on (1) device type (82.3%), (2) mouse click time stamps (70.3%), and (3) geolocation (57.8%). However, our findings also revealed that whether and to what extent a respondent was positively (or negatively) disposed toward paradata use was likely to depend on the type of paradata in question. The three types of paradata addressed in the experimental question appear to have differed in their level of perceived sensitivity or intrusiveness, as was reflected in the varying degrees of favorable attitudes toward their use. Geolocation information, which was the only paradata type that could be considered personal data, appears to have been especially sensitive. Despite these varying degrees of favorable attitudes, we found strong associations between the respondents' attitudes toward each of the three paradata types ([1] vs. [2]: Cramer's $V = .497$, $p < .001$; [1] vs. [3]: Cramer's $V = .461$, $p < .001$; [2] vs [3]: Cramer's $V = .510$, $p < .001$).

As indicated by significant χ^2 tests of independence, the respondents' attitudes toward all three types of paradata varied significantly depending on whether a short definition was provided and a benefit was mentioned. Pair-wise comparisons provided further information on differences between the four conditions. In the case of mouse click time stamps, respondents were significantly less likely to select the "don't know" option when a short definition was provided. Although this was found also for device type and geolocation, differences were not statistically significant. Thus, we found limited support for the assumption that additional information helps respondents to formulate an informed opinion (Hypothesis 1 partly confirmed). Furthermore, respondents were more likely to have a positive attitude toward the use of mouse click time stamps and geolocation information when a short definition was provided (Hypothesis 2 confirmed). This was also the case for information about the type of device used, although differences were not statistically significant. Mentioning benefits for the researchers or the respondents in addition to providing a short definition did not further improve the respondents' attitudes toward paradata use. Moreover, no differences in respondents' attitudes were found depending on whether benefits for researchers or benefits for respondents were mentioned (Hypothesis 3 and Hypothesis 4 not confirmed).

The experimental question varied across the four conditions with respect to the length and complexity of the item wordings. Having found virtually no significant differences in respondents' attitudes in experimental groups (c) and (d), we examined page-level response times as a proxy measure for information processing in order to test the alternative explanation that respondents in these two experimental groups had simply not read and processed the additional information provided. A one-way analysis of variance revealed significant differences in response times between the control group and the experimental groups, $F(3, 2153) = 20.08$, $p < .001$, $\eta^2 = .027$. Pair-wise comparisons of group means using Bonferroni correction revealed that mean response times in the control group (31.6 s) were significantly shorter compared to all three experimental groups ($p < .01$). Mean response times in the first experimental group, (b), where respondents received a short definition, were 35.9 s. This in turn was significantly shorter compared to mean response times in experimental groups (c) and (d), where, in addition to the short definition, either a researcher benefit or a respondent benefit was mentioned. Here, mean response times were 37.5 and 40.5 s, respectively (both significant with $p < .01$). Although the length and complexity of the item wordings was comparable in the experimental groups (c) and (d), the information about a respondent benefit tended to take longer to be processed than the information about a researcher benefit ($p < .10$). In both cases, respondents took note of the benefits mentioned; however, benefits relevant to the respondents themselves appear to have been processed more carefully.

Overall, providing a short definition of the paradata in question seems to have had a greater impact on respondents' attitudes toward the use of those paradata, whereas the impact of additionally making benefits explicit was limited. Although mentioning a benefit for researchers appears to have had no effect, we found some indication that mentioning a benefit for respondents may have had a positive effect on the respondents' attitudes toward paradata use.

Table 3. Logistic Regression on the Likelihood of Having No Opinion About Paradata Use.

	(1) Device Type			(2) Time stamps			(3) Geolocation		
	B	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*
Informed consent									
No definition/no benefit (ref.)									
Definition/no benefit	−0.29	0.21	0.75	−0.57**	0.18	0.57	−0.48*	0.20	0.62
Definition/researcher benefit	0.19	0.19	1.21	−0.29	0.17	0.75	−0.03	0.18	0.97
Definition/respondent benefit	−0.28	0.20	0.76	−0.53**	0.18	0.59	0.04	0.18	1.04
Gender (ref.: men)	0.31*	0.15	1.36	0.46**	0.13	1.58	0.58***	0.14	1.79
Age									
18–29 (ref.)									
30–49	0.02	0.20	1.02	−0.05	0.19	0.95	0.10	0.20	1.11
50–69	−0.25	0.22	0.78	−0.06	0.20	0.94	0.17	0.21	1.19
Education									
Low (ref.)									
Medium	−0.50**	0.18	0.61	−0.33*	0.16	0.72	−0.59***	0.17	0.56
High	−0.29	0.18	0.75	−0.34*	0.16	0.71	−0.37*	0.16	0.69
Survey assessment from 1 (*not at all*) to 5 (*very*)									
Interesting	−0.48***	0.12	0.62	−0.50***	0.11	0.61	−0.46***	0.11	0.63
Varied	−0.10	0.11	0.91	0.01	0.10	1.01	0.00	0.11	1.00
Important for science	−0.22*	0.09	0.80	−0.24**	0.08	0.78	−0.28**	0.08	0.76
Long	−0.10	0.07	0.91	−0.09	0.06	0.91	−0.10	0.07	0.91
Difficult	0.01*	0.09	1.01	0.02	0.08	1.02	0.08	0.08	1.08
Too personal	0.18*	0.08	1.20	0.07	0.07	1.07	0.11	0.08	1.12
Device type (ref.: desktop)	−0.40*	0.20	0.67	−0.35*	0.18	0.70	−0.18	0.18	0.83
(Constant)	0.94	0.59	2.55	1.13*	0.54	3.08	0.29	0.55	1.33
Model parameters	$\chi^2 = 130.14$, $df = 15$, $p < .001$ $N = 2{,}198$; Pseudo-$R^2 = .12$			$\chi^2 = 132.86$, $df = 15$, $p < .001$ $N = 2{,}198$; Pseudo-$R^2 = .11$			$\chi^2 = 138.29$, $df = 15$, $p < .001$ $N = 2{,}198$; Pseudo-$R^2 = .12$		

Note. *B* = logit coefficient; *SE* = standard error; *OR* = odds ratio.
*$p < .05$. **$p < .01$. ***$p < .001$.

We asked the respondents in an open-ended question for their potential reasons for not consenting to paradata use in web surveys. Responses to this question revealed that issues relating to privacy concerns (35.2%) and data abuse (30.1%) were among the most frequently cited reasons. This is in line with findings from previous studies (Couper & Singer, 2013; Revilla et al., 2019). The share of responses addressing privacy or data abuse issues among all responses to the open-ended question was similar across the conditions for both privacy concerns (control group [a] 35.9% and experimental groups [b] 34.7%, [c] 34.4%, and [d] 35.6%; $\chi^2[3, N = 1{,}370] = .223$, *ns*) and data abuse (control group [a] 29.3% and experimental groups [b] 31.3%, [c] 29.0%, and [d] 31.1%; $\chi^2[3, N = 1{,}370] = .671$, *ns*).

Table 3 presents the results of a binary logistic regression on the likelihood of respondents indicating that they had no opinion about paradata use. In this analysis, we wanted to address two questions: First, whether the provision of additional information helped respondents to formulate an informed opinion; second, what respondent-related characteristics had an influence on forming a positive or negative attitude toward paradata use as opposed to indicating not having an opinion on that issue.

Table 4. Logistic Regression on the Likelihood of Having a Positive Attitude toward Paradata Use.

	(1) Device Type			(2) Time stamps			(3) Geolocation		
	B	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*
Informed consent									
No definition/no benefit (ref.)									
Definition/no benefit	0.31	0.17	1.36	0.47**	0.14	1.60	0.48***	0.13	1.62
Definition/researcher benefit	−0.07	0.16	0.93	0.24	0.14	1.27	0.35**	0.13	1.42
Definition/respondent benefit	0.46**	0.17	1.58	0.32*	0.13	1.38	0.30*	0.13	1.35
Gender (ref.: men)	−0.10	0.12	0.90	−0.10	0.10	0.90	−0.44***	0.09	0.65
Age									
18–29 (ref.)									
30–49	0.10	0.16	1.10	0.30*	0.14	1.35	0.35**	0.13	1.41
50–69	0.27	0.18	1.31	0.22	0.15	1.24	0.61***	0.14	1.85
Education									
Low (ref.)									
Medium	0.20	0.15	1.22	0.09	0.12	1.10	−0.15	0.12	0.86
High	−0.03	0.15	0.98	−0.06	0.12	0.94	−0.48***	0.12	0.62
Survey assessment from 1 (*not at all*) to 5 (*very*)									
Interesting	0.39***	0.10	1.48	0.36***	0.08	1.44	0.36***	0.08	1.43
Varied	0.14	0.09	1.15	0.10	0.08	1.11	0.12	0.08	1.12
Important for science	0.28***	0.07	1.32	0.26***	0.06	1.30	0.28***	0.06	1.33
Long	−0.01	0.06	0.99	−0.06	0.05	0.94	−0.15**	0.05	0.86
Difficult	0.01	0.07	1.01	0.00	0.06	1.00	0.12*	0.06	1.13
Too personal	−0.22**	0.07	0.80	−0.12*	0.06	0.88	−0.12*	0.05	0.89
Device type (ref.: desktop)	0.15	0.15	1.16	0.17	0.13	1.19	−0.19	0.12	0.83
(Constant)	−1.36**	0.49	0.26	−1.91***	0.43	0.15	−2.00***	0.42	0.13
Model parameters	$\chi^2 = 186.43$, $df = 15$, $p < .001$ $N = 2{,}198$; Pseudo-$R^2 = .14$			$\chi^2 = 199.87$, $df = 15$, $p < .001$ $N = 2{,}198$; Pseudo-$R^2 = .12$			$\chi^2 = 309.80$, $df = 15$, $p < .001$ $N = 2{,}198$; Pseudo-$R^2 = .18$		

Note. *B* = logit coefficient, *SE* = standard error, *OR* = odds ratio.
*$p < .05$. **$p < .01$. ***$p < .001$.

Our findings revealed that the likelihood of respondents selecting the "don't know" option, and thus indicating that they had no opinion on the use of mouse click time stamps and geolocation information, was significantly lower when a short definition was provided. These results support the assumption that additional information helps respondents formulate an informed opinion on the use of these paradata. Considering several respondent-related characteristics, the following effects were consistently found across the different paradata types: Female respondents were more likely to indicate that they had no opinion concerning the use of web paradata. By contrast, the likelihood of taking a clear position in favor of or against sharing paradata (mouse click time stamps and geolocation) was higher among respondents with a medium or high level of education compared to lower educated respondents. When we examined different aspects of the respondents' assessments of the survey, which we measured on a 5-point fully labeled scale ranging from 1 (*not at all*) to 5 (*very*), we found for type of device used and mouse click time stamps that respondents who considered the questionnaire "interesting" or "important for science," or who answered the survey on a smartphone rather than a desktop computer or tablet, were more likely to take a clear position in favor of or against sharing these paradata.

Table 4 presents the results of a further binary logistic regression on the likelihood of respondents having a positive attitude toward paradata use. This analysis was aimed at identifying factors that

contribute to a positive attitude toward paradata use because we consider these influencing factors to be relevant for researchers who want to encourage paradata consent in their web surveys. In this analysis, respondents with a negative attitude or with no opinion were combined. This decision was based on the conservative assumption that respondents who have an ambivalent or indifferent opinion on the issue of paradata use will be more likely to oppose paradata sharing in cases of doubt.

In line with our descriptive findings, the likelihood of respondents having a positive attitude toward the use of information on device type, mouse click time stamps, and geolocation was significantly greater when a short definition or a short definition in combination with a respondent benefit was provided compared to when no additional information was given. With regard to respondents' sociodemographic characteristics (i.e., gender, age, and education), significant effects were found only for geolocation information. The likelihood of having a negative attitude toward the use of geolocation information was greater among female respondents than among male respondents and among respondents with a higher level of education than among lower educated respondents. Although we had no ex ante assumption regarding age, it was nevertheless surprising that respondents aged 30 years and older were more likely to have a positive attitude toward the use of geolocation information than the youngest age group of 18- to 29-year-old respondents.

Moreover, some aspects of the respondents' assessments of the survey were relevant for forming a positive attitude toward all three paradata types. Respondents who considered the questionnaire "interesting" and "important for science" were more likely to have a favorable attitude toward the use of different types of web paradata. By contrast, respondents who perceived the questionnaire as "too personal" were less likely to be positively disposed toward paradata use. Thus, in line with our expectation, an overall positive assessment of the survey was related to a positive attitude toward web paradata use (Hypothesis 5 confirmed). Also in line with previous findings, a high level of perceived intrusiveness resulted in more skeptical attitudes toward paradata use. The type of device respondents used to answer the survey (i.e., desktop computer/tablet vs. smartphone) had no effect on their likelihood of having a positive attitude toward paradata use (Hypothesis 6 not confirmed).[1]

Discussion

Because a positive attitude toward paradata use is considered to be a prerequisite for respondents' willingness to share web paradata and consent to their collection and use, the present study aimed to identify factors that positively affect respondents' attitudes in this regard. Drawing on the reasoned action framework proposed by Fishbein and Ajzen (2010), we examined different ways of positively influencing respondents' cost–benefit balancing in favor of paradata use. Our results confirm that providing respondents with additional information about web paradata, and enabling them to better understand what paradata are collected, helps them to form an opinion about the use of these paradata and contributes to a more positive attitude toward web paradata use. In this regard, we argue that clarifying information puts respondents in a better position to assess the positive and negative consequences of paradata use and helps mitigate anticipated costs of web paradata sharing. Similarly, explicitly mentioning the benefits of paradata use in order to exemplify the reasons why paradata are gathered is likely to outweigh possible costs. As the provision of additional information such as definitions or benefits of paradata is something that is under the researcher's control, it is a viable way to influence respondents' attitudes and create good prerequisites for consent to web paradata use. However, when respondents are given information about benefits of paradata use, they must recognize and acknowledge them as beneficial. Otherwise, mentioning benefits is of no use for improving respondents' attitudes toward web paradata use.

Although, in the present study, repondents' attitudes toward web paradata use were found to be rather positive, they varied across the different types of paradata and between respondents. Varying degrees of favorable attitudes toward web paradata use depending on the different types of paradata suggest, for instance, that information on the respondent's location at the time of the survey appears to have been considered more sensitive than information on the type of device used to complete the survey. Irrespective of the type of paradata, the likelihood of respondents having a positive attitude toward paradata use depended on whether they had a positive attitude toward the survey. This finding highlights the importance of providing the respondents with well-designed questionnaires and positive survey experiences, which should not only increase overall data quality in web surveys but also enhance rates of consent to web paradata use.

The present study has a number of limitations that could yield opportunities for future research. First, as our findings are based on data from a nonprobability, opt-in online panel, this might impair their generalizability. This is because opt-in online panel members can be considered to be generally more cooperative and presumably more favorably disposed toward sharing paradata in web surveys. For this reason, we used an experimental setup to measure our variables of interest and control for unobserved heterogeneity. Even among panelists with on average fairly positive views toward paradata, considerable variations in their views were found depending on the type of paradata. However, replication of the present study based on other samples would be desirable.

The second limitation of our study is that we selected the three types of paradata based on the fact that they are frequently used for methodological purposes. However, the observed varying degrees of favorable attitudes toward different paradata types suggest that it should be carefully decided what types of paradata are explicitly mentioned when explaining "paradata" as a whole. Future research in this regard could focus on whether it is preferable to mention just one type of paradata or to present several paradata types together. If the latter proves to be the case, insights are needed into the types of paradata that should be presented together in order to adequately inform respondents about what paradata are and why they are used.

Third, the benefits that were mentioned to respondents were benefits that we presumed were of relevance either to researchers or to respondents. However, we did not systematically test in advance whether respondents understood the benefits mentioned and acknowledged them as beneficial. Thus, another open research question is what benefits should be made explicit to respondents and how they should be worded so that they are actually recognized and acknowledged as beneficial by the respondents.

The fourth limitation of our study is that we asked respondents about their attitudes toward web paradata use and not for their consent to web paradata use. According to the reasoned action framework (Fishbein & Ajzen, 2010), however, the attitude toward a behavior is only one of three predictors of behavioral intentions and behaviors. Moreover, the relative importance, or weight, of the different predictors can vary from one individual to another, and a variety of individual, social, or informational background factors may come into play, thereby further adding to the complexity of explaining and predicting behavioral intentions and behaviors. Thus, our study lays the ground for future research that investigates the link between positive attitudes toward and actual consent to web paradata use. In our view, this line of research should not only focus on increasing rates of consent to paradata use but should also investigate the factors influencing the respondents' beliefs about the positive and negative consequences of the consent decision. With the knowledge thus obtained, survey researchers will be better able to encourage respondent consent in a systematic and theoretically grounded way. In this context, our study provides valuable insights into the importance of adequately informing respondents about what paradata are and why they are being used, in order to improve their attitudes toward web paradata use.

Appendix A

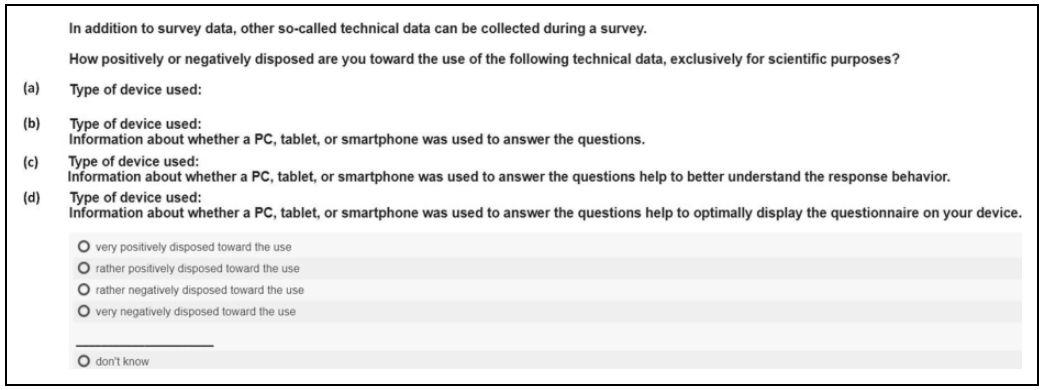

Figure A1. Paradata description texts in the control group (a) and the three experimental groups (b), (c), and (d) using the item on "type of device used" as an example.

Data Availability

The data set generated and analyzed during the current study is available on request from the corresponding author. Email: tanja.kunz@gesis.org

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Software Information

All analyses in the present study were conducted using IBM SPSS Statistics version 24.0.

Note

1. As a robustness check, we reran these models and included additional variables relating to the respondents' prior survey experience, the duration of their panel membership, and their experience in using computers and the Internet. These variables neither had a significant effect on the respondents' attitudes toward paradata use nor did their inclusion change the substantive findings obtained on the effects of our variables of interest. We interpret the results of the robustness checks as support for our conclusions.

References

ADM e.V., ASI e.V., BVM e.V., & DGOF e.V. (2007). *Richtlinie für Online-Befragungen* [*Guideline for online surveys*]. Retrieved October 25, 2018, from http://rat-marktforschung.de/fileadmin/user_upload/pdf/R08_RDMS.pdf

American Sociological Association. (2018). *Code of ethics*. Retrieved October 25, 2018, from http://www.asanet.org/sites/default/files/asa_code_of_ethics-june2018.pdf

Callegaro, M. (2013). Paradata in web surveys. In F. Kreuter (Ed.), *Improving surveys with paradata. Analytic uses of process information* (pp. 261–279). Hoboken, NJ: Wiley.

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*, 1008–1032.

Conrad, F. G., Tourangeau, R., Couper, M., & Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods*, *11*, 45–61.

Couper, M. P. (1998). Measuring survey quality in a CASIC environment. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, 9–13 August, Dallas, TX.

Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. A total survey error perspective. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 133–154). Hoboken, NJ: Wiley.

Couper, M. P., & Singer, E. (2013). Informed consent for web paradata use. *Survey Research Methods*, *7*, 57–67.

ESOMAR/GRBN. (2015). *Guideline on Online Research*. Retrieved October 25, 2018, from https://www.esomar.org/what-we-do/code-guidelines/esomargrbn-online-research-guideline

ESOMAR/GRBN. (2017). *Guideline on Mobile Research*. Retrieved October 25, 2018, from https://www.esomar.org/what-we-do/code-guidelines/esomargrbn-guideline-on-mobile-research

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior. The reasoned action approach*. New York, NY: Taylor and Francis Group.

Gummer, T., Quoß, F., & Roßmann, J. (2019). Does increasing mobile device coverage reduce heterogeneity in completing web surveys on smartphones? *Social Science Computer Review*, *37*, 371–384. doi:10.1177/0894439318766836

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, *21*, 360–373.

Kaczmirek, L., & Neubarth, W. (2007). Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Paradaten evaluieren [Non-reactive data collection: Evaluating participation behavior in surveys with paradata]. In M. Welker & O. Wenzel (Eds.), *Online-Forschung 2007. Grundlagen und Fallstudien* [*Online research 2007. Fundamentals and case studies*] (pp. 293–311). Cologne, Germany: Herbert von Halem Verlag.

Keusch, F., Antoun, C., Couper, M. P., Kreuter, F., & Struminskaya, B. (2017, May). *Willingness to participate in passive mobile data collection.* Paper presented at the 72nd AAPOR Annual Conference, New Orleans, LA.

Kreuter, F. (2013). Improving surveys with paradata: Introduction. In F. Kreuter (Ed.), *Improving surveys with paradata* (pp. 1–9). Hoboken, NJ: John Wiley & Sons, Inc.

Kunz, T. (2015). *Rating scales in web surveys. A test of new drag-and-drop rating procedures* (Doctoral dissertation). Darmstadt University of Technology, Darmstadt, Germany. Retrieved July 16 from http://tuprints.ulb.tu-darmstadt.de/id/eprint/5151

Mavoa, S., Oliver, M., Witten, K., & Badland, H. M. (2011). Linking GPS and travel diary data using sequence alignment in a study of children's independent mobility. *International Journal of Health Geographics*, *10*, 1–10.

McClain, C. A., Couper, M. P., Hupp, A. L., Keusch, F., Peterson, G., Piskorowski, A. D., & West, B. T. (2019). A typology of web survey paradata for assessing total survey error. *Social Science Computer Review*, *37*, 196–213. doi:10.1177/0894439318759670

Revilla, M. A., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses*, *13*, 223–252. doi:10.12758/mda.2018.01

Revilla, M. A., & Ochoa, C. (2015). What are the links in web survey among response time, quality, and auto-evaluation of the efforts done? *Social Science Computer Review*, *33*, 97–114.

Sakshaug, J. W., & Kreuter, F. (2014). The effect of benefit wording on consent to link survey and administrative records in a web survey. *Public Opinion Quarterly*, *78*, 166–176.

Schlosser, S., & Höhne, J. K. (2018). *ECSP – Embedded Client Side Paradata*. Retrieved July 16, 2018, from https://zenodo.org/record/1218941#.WuCaYW6FMWq

Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, *26*, 1488–1495.

Struminskaya, B., Toepoel, V., Lugtig, P., Schouten, B., Haan, M., & Luiten, A. (2018, May). *Willingness to collect smartphone sensor data in a Dutch probability-based general population panel.* Paper presented at the 73rd AAPOR Conference, Denver, CO.

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K., & Gaia, A. (2018). Collecting multiple data linkage consents in a mixed-mode survey: Evidence from a large-scale longitudinal study in the UK. *Survey Methods: Insights From the Field*. doi:10.13094/SMIF-2018-00004

Toepoel, V., & Lugtig, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of internet users. *Social Science Computer Review*, *32*, 544–560.

Wenz, A., Jäckle, A., & Couper, M. P. (2017). *Willingness to use mobile technologies for data collection in a probability household panel*. Retrieved October 25, 2018, from https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2017-10.pdf

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*, 51–68.

Author Biographies

Tanja Kunz is a senior researcher at GESIS—Leibniz Institute for the Social Sciences, Germany. Her current research interests include issues of visual design and data quality in web surveys, paradata, and questionnaire design.

Tobias Gummer is a senior researcher and team leader at GESIS—Leibniz Institute for the Social Sciences, Germany. His methodological research interests include survey design, data quality, nonresponse, and correction methods for biases.